Adnexal mass risk assessment: a multivariate index assay for malignancy risk stratification

Zhen Zhang¹, Rowan G Bullock^{*,2} & Herbert Fritsche³¹School of Medicine, Johns Hopkins University, Baltimore, MD 21218, USA²Vermillion, Inc., Research and Development Department, Austin, TX 78738, USA³ASPIRA Labs, Austin, TX 78738, USA*Author for correspondence: rbullock@vermillion.com

Aims: Adnexal mass risk assessment (AMRA) stratifies patients with adnexal masses, identifying the relatively small number of malignancies from benigns which might take a 'watchful waiting' approach. **Methods:** AMRA uses seven biomarkers and derived from women with adnexal masses scheduled for surgery. Estimated clinical performance was calculated using fixed prevalence. **Results:** At 5% prevalence, the high-risk group, 7.9% total, captured 75.9% of invasive malignancies at a positive predictive value of 35.8%. High risk/intermediate risk combined had a sensitivity of 89.7 and 95.6% for pre- and post-menopausal cancers, respectively. The low-risk group, 67.8% total, had a negative predictive value of 99.0%. **Conclusion:** With highly differentiating risk stratification capability across histological subtypes and stages, AMRA is potentially applicable to patients with adnexal masses to assist deciding whether immediate surgery is recommended.

First draft submitted: 7 August 2019; Accepted for publication: 9 September 2019; Published online: 1 October 2019

Keywords: adnexal mass • multivariate index assay • ovarian cancer • watch and wait

Adnexal masses are a common clinical diagnosis in women from the age of adolescence onward, occurring in five to ten percent of women during their lifetimes. Masses are benign in the vast majority of cases [1]. However, the potential risk of malignancy must be considered.

Benign masses may cause problems due to size, proximity to organs and pain or discomfort. Surgical removal of the mass may be recommended. However, some benign masses may be largely asymptomatic and can be monitored. The benefits of this are clear: the avoidance of an invasive surgery and the associated costs, both financial and in terms of recovery time [2].

The current standard of care for a suspected adnexal mass is imaging, typically performed via transvaginal ultrasonography. Imaging may clearly reveal the mass to be benign or malignant based on its physical features, but there are many cases in which the mass is indeterminate [3]. In these cases, in an asymptomatic patient, follow-up imaging after a period of time is recommended in order to determine if the mass is persistent, stable or has resolved. A risk associated with this approach is delay in detection and management of a malignant lesion [4].

Adnexal mass risk assessment (AMRA) is a multivariate index assay of serum biomarkers developed to segregate patients with a suspected adnexal mass into three risks of malignancy categories to assist in the decision on whether immediate surgery is recommended. The objective is to use one cutoff to capture a high percentage of the total cancer cases within a relative small group of high risk patients resulting a clinically actionable positive predictive value and to use a second cutoff to assign a significant portion of the remaining test population into a low-risk (LR) group with a very high negative predictive value.

The ability to develop and evaluate a serum test for the above-described intended use is often hindered by the lack of properly enrolled clinical samples with inclusion/exclusion criteria representative of the targeted test population and the practical difficulty of missing definitive clinical classifications for those who do not have surgery. For the development as well as independent validation of AMRA, biomarker data were retrospectively analyzed from prospectively collected cohorts of patients diagnosed with adnexal masses and for whom all had definitive pathological classifications from surgery. To project the performance of AMRA on its intended population of

patients with adnexal masses prior to the decision of surgery, we assumed that except for the obvious differences in cancer prevalence, the cancer cases in the patient cohorts used in the current study and AMRA's intended population are similar in terms of histologic subtypes, grades, and stages, and if anything else the noncancer benign patients in AMRA's intended population would likely be 'healthier' controls. This allowed us to estimate conservatively important test performance metrics such as positive and negative predictive values of the AMRA high- and low-risk groups, respectively by adjusting for an assumed malignancy prevalence that is typically observed in AMRA's intended populations [8].

Methods

Multi-analyte panel

Seven serum protein analytes were included as input for AMRA development: APO-A1, B2M, CA 125-II, FSH, HE4, TRF and TT. Subsets of these analytes have been used for multivariate index assay (MIA) and multivariate index assay second generation (MIA2G). For all datasets, the analytes were measured on Roche cobas 6000 (Roche Diagnostics Corp., IN, USA) per manufacturer product package inserts.

Sample sets

Data from samples in four separate collections were used in this study. For the derivation of the AMRA algorithm, the training data were from samples that have been previously used for the derivation and validation for the original MIA and the MIAG2 IVDMA [5,6,7] tests ('OVA1 Study'). These samples were originally collected prospectively from 27 Institutional Review Board-approved sites throughout the USA. Inclusion criteria were: women age ≥ 18 years, signed informed consent, agreeable to phlebotomy and documented pelvic mass planned for surgical intervention within 3 months of imaging. A pelvic mass was confirmed by imaging (computed tomography, ultrasonography or MRI) prior to enrollment. Exclusion criteria included a diagnosis of malignancy in the previous 5 years (excepting nonmelanoma skin cancers). With the original use of the sample collection to evaluate the utility of MIA and MIAG2 in referring high-risk (HR) patients to be operated by gynecologic oncologists, the collection excluded patient initially enrolled by a gynecologic oncologist. Menopause was defined as the absence of menses for ≥ 12 months, or age ≥ 50 years for a small number of subjects for whom the menopausal data were missing. Demographic and clinicopathological data were collected on case report forms.

The original prospectively collected sample set represented the actual test population of preoperative risk assessment of adnexal masses. For AMRA algorithm development, however, only a subset (88.36%) of the total samples for whom results of all seven analytes were available, resulting in a slightly over-representation of ovarian cancer cases. Among the 585 samples, there were 284 premenopausal patients including 54 ovarian cancer cases and 230 benign controls, and 301 postmenopausal patients with 124 cases and 177 controls. Table 1 lists in detail relevant demographic and clinicopathological descriptions of the sample sets.

The trained AMRA algorithms were then validated on datasets from the remaining three independent Institutional Review Board-approved specimen cohorts: FHCRC#7788, the OVA500 study and OVA1-PS1-CO4.

For FHCRC#7788, patients were prospectively enrolled at gynecologic oncology and benign gynecologic clinics at the Seattle Cancer Care Alliance and the University of Washington Medical Center from 2012 to 2015. Inclusion criteria included women age ≥ 18 years, signed informed consent and a documented adnexal mass planned for surgery. An adnexal mass was confirmed by imaging (computed tomography, ultrasonography or MRI) prior to enrollment. Exclusion criteria included pelvic surgery within 6 weeks prior to presentation. Demographic, clinical and pathologic information were collected prospectively. Informed consent was provided by all enrolled patients. For validation of AMRA, a case-control set from the FHCRC#7788 cohort was used based on samples with all seven biomarkers.

The OVA500 study and OVA1-PS1-CO4 had the same enrollment and exclusion criteria as the OVA1 study except that OVA1-PS1-CO4 included only subjects not yet referred to a gynecologic oncologist even though for whom surgical intervention had been planned. As a result, the prevalence of ovarian cancer in OVA1-PS1-CO4 is significantly lower than that in OVA500.

All samples in each study with known biomarker values for all seven analytes were used in this analysis. Among the cancer cases, 15 samples (0.074% of total samples) with malignancies not involved with the ovary were removed from analysis.

Table 1. Study demographics and clinicopathologic information according to menopausal stage.

Variables	All subjects			OVA 1 study (training set)			OVA 500 study (validation set)			FHCRC#7788 (validation set)			OVA1-PS1-CO4 (validation set)		
	All (n = 2092)	Pre (n = 1237)	Post (n = 855)	All (n = 574)	Pre (n = 281)	Post (n = 293)	All (n = 511)	Pre (n = 287)	Post (n = 224)	All (n = 218)	Pre (n = 70)	Post (n = 148)	All (n = 789)	Pre (n = 599)	Post (n = 190)
Age (years)															
Mean (SD)	48.0	39.4	60.3	51.0	40.6	61.0	48.6	39.5	60.2	53.6	40.3	59.8	43.8	38.79	59.5
Median	47	41	60	50	42	60	47	41	60	54	42	60	43	40	58
Range (min–max)	18–92	18–60	31–92	18–92	18–58	37–92	18–87	18–60	33–87	18–86	18–59	31–86	18–86	18–58	37–86
Race/ethnicity, n (%)															
White/Caucasian	1566 (74.86)	851 (68.80)	715 (83.63)	451 (78.57)	196 (69.75)	255 (87.03)	358 (70.06)	177 (61.67)	181 (80.80)	179 (82.11)	52 (74.29)	127 (85.81)	578 (73.26)	426 (71.12)	152 (80.00)
Black/African-American	255 (12.19)	189 (15.28)	66 (7.72)	62 (10.80)	45 (16.01)	17 (5.80)	85 (16.63)	58 (20.21)	27 (12.05)	1 (0.46)	0 (0.00)	1 (0.68)	107 (13.56)	86 (14.36)	21 (11.05)
Hispanic/Latino	178 (8.51)	136 (10.99)	42 (4.91)	46 (8.01)	30 (10.68)	16 (5.46)	49 (9.59)	39 (13.59)	10 (4.46)	0 (0.00)	0 (0.00)	0 (0.00)	83 (10.52)	67 (11.19)	16 (8.42)
Asian	38 (1.82)	27 (2.18)	11 (1.29)	9 (1.57)	6 (2.14)	3 (1.02)	13 (2.54)	8 (2.79)	5 (2.23)	4 (1.83)	1 (1.43)	3 (2.03)	12 (1.52)	12 (2.00)	0 (0.00)
Native American or Alaska Native	9 (0.43)	5 (0.40)	4 (0.47)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	5 (2.29)	2 (2.86)	3 (2.03)	4 (0.51)	3 (0.50)	1 (0.53)
Native Hawaiian or Pacific Islander	6 (0.29)	6 (0.49)	0 (0.00)	1 (0.17)	1 (0.36)	0 (0.00)	1 (0.20)	1 (0.35)	0 (0.00)	2 (0.92)	2 (2.86)	0 (0.00)	2 (0.25)	2 (0.33)	0 (0.00)
Middle Eastern	3 (0.14)	2 (0.16)	1 (0.12)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	2 (0.92)	1 (1.43)	1 (0.68)	1 (0.13)	1 (0.17)	0 (0.00)
Other	22 (1.05)	14 (1.13)	8 (0.94)	5 (0.87)	3 (1.07)	2 (0.68)	5 (0.98)	4 (1.39)	1 (0.45)	10 (4.59)	5 (7.14)	5 (3.38)	2 (0.25)	2 (0.33)	0 (0.00)
Decline to answer/unknown	15 (0.72)	7 (0.57)	8 (0.94)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	11 (5.05)	6 (8.57)	5 (3.38)	0 (0.00)	0 (0.00)	0 (0.00)
Pathology diagnosis, n (%)															
Benign ovarian conditions	1688 (79.81)	1107 (88.92)	576 (66.59)	407 (69.57)	230 (80.99)	177 (58.80)	415 (79.96)	255 (87.33)	160 (70.48)	124 (56.88)	47 (67.14)	77 (52.03)	742 (93.57)	575 (95.99)	162 (85.71)
EOC	260 (12.29)	66 (5.30)	194 (22.43)	106 (18.12)	28 (9.86)	78 (25.91)	64 (12.33)	19 (6.51)	45 (19.82)	61 (27.98)	8 (11.43)	53 (35.81)	29 (3.66)	11 (1.84)	18 (9.52)
Nonepithelial primary ovarian cancer	29 (1.37)	17 (1.37)	12 (1.39)	12 (2.05)	4 (1.41)	8 (2.66)	5 (0.96)	5 (1.71)	0 (0.00)	5 (2.29)	2 (2.86)	3 (2.03)	7 (0.88)	6 (1.00)	1 (0.53)
Low malignant potential (borderline)	72 (3.40)	31 (2.49)	41 (4.74)	29 (4.96)	11 (3.87)	18 (5.98)	17 (3.28)	5 (1.71)	12 (5.29)	17 (7.80)	9 (12.86)	8 (5.41)	9 (1.13)	6 (1.00)	3 (1.59)
Nonprimary ovarian malignancies with involvement of the ovaries	28 (1.32)	10 (0.80)	18 (2.08)	20 (3.42)	8 (2.82)	12 (3.99)	6 (1.16)	2 (0.68)	4 (1.76)	0 (0.00)	0 (0.00)	0 (0.00)	2 (0.25)	0 (0.00)	2 (1.06)
EOC: Epithelial ovarian cancer.															

Table 1. Study demographics and clinicopathologic information according to menopausal stage (cont.).

Variables	All subjects			OVA1 study (training set)			OVA500 study (validation set)			FHCRC#7788 (validation set)			OVA1-PS1-CO4 (validation set)		
	All (n = 2092)	Pre (n = 1237)	Post (n = 855)	All (n = 574)	Pre (n = 281)	Post (n = 293)	All (n = 511)	Pre (n = 287)	Post (n = 224)	All (n = 218)	Pre (n = 70)	Post (n = 148)	All (n = 789)	Pre (n = 599)	Post (n = 190)
Stage (EOC and other primary), n (%)															
Stage I	87 (30.10)	29 (34.94)	58 (28.16)	36 (30.51)	14 (43.75)	22 (25.58)	30 (43.48)	9 (37.50)	21 (46.67)	13 (19.70)	2 (20.00)	11 (19.64)	8 (22.22)	4 (23.53)	4 (21.05)
Stage II	33 (11.42)	12 (14.46)	21 (10.19)	18 (15.25)	8 (25.00)	10 (11.63)	7 (10.14)	2 (8.33)	5 (11.11)	6 (9.09)	1 (10.00)	5 (8.93)	2 (5.56)	1 (5.88)	1 (5.26)
Stage III	128 (44.29)	28 (33.73)	100 (48.54)	57 (48.31)	9 (28.13)	48 (55.81)	26 (37.68)	10 (41.67)	16 (35.56)	37 (56.06)	7 (70.00)	30 (53.57)	8 (22.22)	2 (11.76)	6 (31.58)
Stage IV	23 (7.96)	4 (4.82)	19 (9.22)	4 (3.39)	0 (0.00)	4 (4.65)	6 (8.70)	3 (12.50)	3 (6.67)	10 (15.15)	0 (0.00)	10 (17.86)	3 (8.33)	1 (5.88)	2 (10.53)
Not given	18 (6.23)	10 (12.05)	8 (3.88)	3 (2.54)	1 (3.13)	2 (2.33)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	15 (41.67)	9 (52.94)	6 (31.58)
Grade (EOC and other primary), n (%)															
1	47 (16.26)	22 (26.51)	25 (12.14)	13 (11.02)	9 (28.13)	4 (4.65)	9 (13.04)	1 (4.17)	8 (17.78)	16 (24.24)	8 (80.00)	8 (14.29)	9 (25.00)	4 (23.53)	5 (26.32)
2	48 (16.61)	17 (20.48)	31 (15.05)	27 (22.88)	10 (31.25)	17 (19.77)	12 (17.39)	4 (16.67)	8 (17.78)	5 (7.58)	1 (10.00)	4 (7.14)	4 (11.11)	2 (11.76)	2 (10.53)
3	178 (61.59)	40 (48.19)	138 (66.99)	70 (59.32)	13 (40.63)	57 (66.28)	43 (62.32)	16 (66.67)	27 (60.00)	49 (74.24)	5 (50.00)	44 (78.57)	16 (44.44)	6 (35.29)	10 (52.63)
Not given/other	22 (7.61)	8 (9.64)	14 (6.80)	8 (6.78)	0 (0.00)	8 (9.30)	5 (7.25)	3 (12.50)	2 (4.44)	2 (3.03)	0 (0.00)	2 (3.57)	7 (19.44)	5 (29.41)	2 (10.53)
Primary histology (EOC and other primary), n (%)															
Epithelial:															
- Serous	133 (46.02)	26 (31.33)	107 (51.94)	61 (51.69)	10 (31.25)	51 (59.30)	24 (34.78)	8 (33.33)	16 (35.56)	38 (57.58)	4 (40.00)	34 (60.71)	10 (27.78)	4 (23.53)	6 (31.58)
- Endometriod	39 (13.49)	15 (18.07)	24 (11.65)	11 (9.32)	7 (21.88)	4 (4.65)	13 (18.84)	5 (20.83)	8 (17.78)	10 (15.15)	1 (10.00)	9 (16.07)	5 (13.89)	2 (11.76)	3 (15.79)
- Mucinous	30 (10.38)	11 (13.25)	19 (9.22)	9 (7.63)	5 (15.63)	4 (4.65)	12 (17.39)	2 (8.33)	10 (22.22)	4 (6.06)	0 (0.00)	2 (3.57)	5 (13.89)	2 (11.76)	3 (15.79)
- Clear cell	22 (7.61)	6 (7.23)	16 (7.77)	9 (7.63)	4 (12.50)	5 (5.81)	5 (7.25)	1 (4.17)	4 (8.89)	4 (6.06)	0 (0.00)	4 (7.14)	4 (11.11)	1 (5.88)	3 (15.79)
- Transitional	3 (1.04)	1 (1.20)	2 (0.97)	2 (1.69)	0 (0.00)	2 (2.33)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (2.78)	1 (5.88)	0 (0.00)
- Carcinosarcoma	10 (3.46)	2 (2.41)	8 (3.88)	5 (4.24)	0 (0.00)	5 (5.81)	1 (1.45)	1 (4.17)	0 (0.00)	3 (4.55)	1 (10.00)	2 (3.57)	1 (2.78)	0 (0.00)	1 (5.26)
- Mixed	10 (3.46)	3 (3.61)	7 (3.40)	3 (2.54)	2 (6.25)	1 (1.16)	2 (2.90)	0 (0.00)	2 (4.44)	2 (3.03)	0 (0.00)	2 (3.57)	3 (8.33)	1 (5.88)	2 (10.53)
- Undifferentiated	3 (1.04)	0 (0.00)	3 (1.46)	3 (2.54)	0 (0.00)	3 (3.49)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
- Other	10 (3.46)	2 (2.41)	8 (3.88)	3 (2.54)	0 (0.00)	3 (3.49)	7 (10.14)	2 (8.33)	5 (11.11)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Nonepithelial:															
- Sex cord stromal	9 (3.11)	4 (4.82)	5 (2.43)	7 (5.93)	3 (9.38)	4 (4.65)	0 (0.00)	0 (0.00)	0 (0.00)	1 (1.52)	0 (0.00)	1 (1.79)	1 (2.78)	1 (5.88)	0 (0.00)
- Germ cell	4 (1.38)	2 (2.41)	2 (0.97)	3 (2.54)	1 (3.13)	2 (2.33)	0 (0.00)	0 (0.00)	0 (0.00)	1 (1.52)	1 (10.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
- Other	16 (5.54)	11 (13.25)	5 (2.43)	2 (1.69)	0 (0.00)	2 (2.33)	5 (7.25)	5 (20.83)	0 (0.00)	3 (4.55)	1 (10.00)	2 (3.57)	6 (16.67)	5 (29.41)	1 (5.26)
EOC: Epithelial ovarian cancer.															

Model derivation

AMRA was developed for the intended utility of using two cutoffs to identify a small group of HR patients that captures a large portion of the cancer cases and a relatively large groups of LR patients with a high negative predictive value. During model derivation, the desired performance characteristics, in particular, to have an improved sensitivity at very high specificity, was translated and implemented numerically and computationally to influence the derivation and selection of the final models. To assure statistical stability of results, extensive statistical resampling approaches, such as bootstrap sampling of data points within the training sample set and random selection of subset of input analytes were extensively used.

Two separate predictive models (algorithms) were derived for the pre- and post-menopausal patient populations, respectively. For each algorithm, the training dataset was also used to determine a ‘rule-in’ cutoff to identify HR group and a ‘rule-out’ cutoff to identify relatively LR group. The samples between the two cutoffs were classified as intermediate risk (IR). The selection of cutoffs was driven by the desired performance characteristics based on consensus from clinicians. For example, tradeoff between sensitivity and having a smaller proportion of patients in HR group (and the corresponding positive predictive value [PPV]) determined the rule-in cutoff; and similarly, balance between a required high negative predictive value (NPV) and having a sufficiently large LR group helped to select the rule-out cutoff.

Performance evaluation

The derived AMRA algorithms and the fixed cutoffs were evaluated on the training set and the three independent validation sets individually and combined. Area under curves from receiver-operating characteristic curve analysis were used to assess the overall performance and to compare with CA125. Sensitivity for the rule-in cutoff (proportion of total cases captured by the HR group) and specificity for the rule-out cutoff (proportion of total benigns in the LR group) were estimated. In addition, positive likelihood ratios (LR+) and negative likelihood ratios (LR-), which are not dependent on prevalence, were also estimated for the HR and LR groups, respectively, and used to provide approximated assessment on changes in post-test probability of cancer from pretest probability.

The samples used for the current study were all from what were originally prospectively collected cohorts. However, with the retrospective use of only samples for whom results of all seven biomarkers were available, the proportion of malignant cases and benign controls were somewhat skewed. The reasons for samples with missing biomarker values were mostly due to specimen volume and other study-dependent decision on data generation and were not directly related to patient clinicopathological phenotypes. As previously discussed, it seemed reasonable to assume that the cancer cases and the benign controls of the individual sample sets were approximate representations of the cases and controls, respectively, in the AMRA test’s intended test population except for the obvious differences in malignancy prevalence. To project the performance of the AMRA algorithms in its intended population, we assumed a pretest prevalence of 5 and 10% in pre- and post-menopausal test populations, respectively. The projected distribution of benign, low-malignant potential tumors (LMPs), stage I/II cases and stage III/IV cases among the three AMRA risk classification groups were estimated after adjustment for the assumed prevalence. Based on such adjustments, percentage of total test population, post-test cancer prevalence was estimated for the AMRA risk groups, including PPVs and NPVs for the HR and LR groups, respectively.

Results

Figure 1 shows the receiver operating characteristic (ROC) curves of AMRA with comparison to CA125 on pre- and post-menopausal patients in the training set, and the three validation sets. AMRA outperformed CA125 across all sample sets, especially in the part of ROC curve representing sensitivity at very high specificity (leftmost of the plot) that determines the performance characteristics of the rule-in (HR) group. In particular, the improvement among premenopausal patients was statistically significant for all ovarian cancer (p-values = 0.01, 0.06, 0.05 and <0.01) as well as for stage I/II invasive ovarian cancer (p-values = 0.01, 0.07, <0.01 and <0.01; Figure 1C) for the training set (OVA1), and the three validation sets (OVA500, FHCRC#7788 and OVA1-PS1-CO4), respectively.

The AMRA rule-in and rule-out cutoff values, at 14.0 IU and 5.0 IU, respectively, for the premenopausal model, and 10.0 IU and 5.0 IU for the postmenopausal model, were established using the training sample set and then fixed in validation. Using the rule-in and rule-out cutoffs, the samples in the three validation sets were combined to estimate and plot the prevalence-adjusted cancer/benign distributions among the three AMRA risk groups (Figure 2). The raw and prevalence-adjusted proportion of cancer and benign patients within each risk groups are listed in Tables 2 and 3. In Table 4, both prevalence-independent performance metrics such as sensitivity, specificity,

Figure 1. Comparison of ROC curves between adnexal mass risk assessment and CA125 on individual sample sets. (A) (left): premenopausal, (B) (middle): postmenopausal, (C) (right): premenopausal, stage I/II invasive cancer and benign adnexal masses only. Training set: OVA1 Study, Validation sets: OVA500 Study, FHCRC #7788 and OVA1-PS1-CO4. AMRA: Adnexal mass risk assessment; AUC: Area under curve; ROC: Receiver operating characteristic.

Figure 2. Bar charts illustrate the distributions of benign, low-malignant potential tumor stage I/II and stage III/IV patients in adnexal mass risk assessment risk groups. Figure 3A (top): premenopausal, Figure 3B (bottom): postmenopausal. Left: training set (OVA1 Study), Right: combined validation Sets (OVA500 Study, FHCRC #7788 and OVA1-PS1-CO4). Exact numbers are listed under 'Estimates Adjusted for Assumed Pretest Prevalence' in Tables 2 & 3. AMRA: Adnexal mass risk assessment; HR: High risk; IR: Intermediate risk; LR: Lower risk.

Table 2. Distribution of benign, low-malignant potential tumor/early stage, late stage cancer in premenopausal adnexal mass risk assessment risk groups (actual and projected based on assumed prevalence).

Study	AMRA risk group	Path	n	Within risk group (%)	Estimates adjusted for assumed pretest prevalence at 5%			
					Adj. percentage within risk group	Percentage of total population	Prevalence (%) in risk group	Prevalence (%) in risk group excl. LMP
Training set: OVA1 study set	HR	Benign	14	28.6	62.8	9.2	37.2	30.9
		LMP	6	12.2	6.4			
		OC Early	14	28.6	14.9			
		OC Late	15	30.6	16.0			
	IR	Benign	74	89.2	97.2	31.4	2.8	2.2
		LMP	2	2.4	0.6			
		OC Early	5	6.0	1.6			
		OC Late	2	2.4	0.6			
	LR	Benign	142	95.3	98.8	59.3	1.2	0.7
		LMP	3	2.0	0.5			
		OC Early	3	2.0	0.5			
		OC Late	1	0.7	0.2			
Validation sets: OVA500, FHCRC #7788, OVA1-PS1-CO4	HR	Benign	42	44.7	57.7	7.9	42.3	35.8
		LMP	8	8.5	6.5			
		OC Early	13	13.8	10.6			
		OC Late	31	33.0	25.2			
	IR	Benign	216	93.5	96.0	24.3	4.0	2.1
		LMP	7	3.0	1.8			
		OC Early	3	1.3	0.8			
		OC Late	5	2.2	1.3			
	LR	Benign	620	98.3	99.0	67.8	1.0	0.6
		LMP	5	0.8	0.5			
		OC Early	3	0.5	0.3			
		OC Late	3	0.5	0.3			

Adj.: Adjusted; AMRA: Adnexal mass risk assessment; excl: Excluding; HR: High risk; IR: Intermediate risk; LMP: Low-malignant potential tumor; LR: Lower risk; OC: Ovarian cancer.

and LR+ and LR-, and the prevalence-adjusted estimates such as percentage of test population, PPV and NPV are provided for the individual and meaningful combinations of risk groups. Finally, for easy visualization, [Figure 3](#) plots the prevalence-adjusted projected post-test cancer probabilities of the AMRA risk groups superimposed with an interpolation curve by logistic regression.

AMRA demonstrated sharp differentiation of post-test cancer probabilities across the three risk groups that should be clinically actionable ([Figure 3](#) with numerical results in [Tables 2 & 3](#)). For both pre- and post-menopausal population, the relatively small HR group (7.9 and 10.6% of test population, respectively) captured over two-thirds of the total cancer cases with PPVs at 42.3 and 66.1%, respectively. When LMPs were excluded, the sensitivity of HR both increased to 75% or above. The LR group identified a significant portion of the test population (67.8 and 52.7% for pre- and post-menopausal, respectively) with an NPV close to 99%. Post-test cancer prevalence in the remaining IR group patients was lower than the assumed pretest prevalence (4.0 and 6.3% for all cancer, or 2.1 and 4.8% excluding LMPs, for pre- and post-menopausal, respectively). [Table 4](#) also listed the sensitivity and PPV of the combined group of HR and IR to be at 85.9 and 13.3%, respectively for premenopausal patients and 93.1 and 19.7%, respectively, for postmenopausal patients. The sensitivities for invasive cancer only were even higher.

Table 3. Distribution of benign, low-malignant potential tumor/early stage, late stage cancer in postmenopausal adnexal mass risk assessment risk groups (actual and projected based on assumed prevalence).

Study	AMRA risk group	Path	n	Within risk group (%)	Estimates adjusted for assumed pretest prevalence at 10%			
					Adj. percentage within risk group	Percentage of total population	Prevalence (%) in risk group	Prevalence (%) in risk group excl. LMP
Training set: OVA1 study set	HR	Benign	25	21.6	61.8	20.6	38.2	35.6
		LMP	6	5.2	2.5			
		OC Early	24	20.7	10.1			
		OC Late	61	52.6	25.6			
	IR	Benign	89	80.9	96.2	47.1	3.8	2.2
		LMP	9	8.2	1.6			
		OC Early	7	6.4	1.3			
		OC Late	5	4.5	0.9			
	LR	Benign	63	94.0	98.9	32.4	1.1	0.3
		LMP	3	4.5	0.8			
		OC Early	1	1.5	0.3			
		OC Late						
Validation sets: OVA500, FHCRC #7788, OVA1-PS1-CO4	HR	Benign	16	12.6	33.9	10.6	66.1	60.8
		LMP	9	7.1	5.4			
		OC Early	36	28.3	21.5			
		OC Late	66	52.0	39.3			
	IR	Benign	154	80.6	93.7	36.7	6.3	4.8
		LMP	9	4.7	1.5			
		OC Early	10	5.2	1.7			
		OC Late	18	9.4	3.1			
	LR	Benign	233	95.5	98.7	52.7	1.3	0.7
		LMP	5	2.0	0.6			
		OC Early	2	0.8	0.2			
		OC Late	4	1.6	0.5			

Adj: Adjusted; AMRA: Adnexal mass risk assessment; excl: Excluding; HR: High risk; IR: Intermediate risk; LMP: Low-malignant potential tumor; LR: Lower risk; OC: Ovarian cancer.

Discussion

AMRA was developed to segregate patients into three groups with highly differentiating post-test cancer probabilities. Under the assumption that the ovarian cancer cases would be similar in terms of histologic subtypes, grades and stages in patients diagnosed with adnexal masses with or without planned surgery, we were able to estimate PPVs and NPVs using the validation datasets individually (Figure 2) [8]. Adjusting for assumed pretest cancer prevalence, we were able to statistically project AMRA potential performance for its intended use in the decision of whether immediate surgery is recommended for the individual validation sets. Finally, to improve statistical stability of the estimated results, the overall validation results were further estimated using the combined validation data, adjusted to the assumed pretest prevalence (Figure 3 & Tables 2–4).

The projected post-test cancer probabilities among the AMRA risk groups, such as PPV for the HR group, and NPV for the LR group, are dependent on the assumed pretest cancer prevalence. The ability of AMRA to segregate patients into highly differential and clinically meaningful risk groups is however further supported by the estimated positive and negative likelihood ratios (LR+ and LR-) which are prevalence-independent measures of how a positive or negative test result might alter the probability of disease. For example, using a simplified interpretation of LRs suggested by Steven McGee [9] the estimated LR+ of > 10.0 for the postmenopausal AMRA HR group in the combined validation samples indicates a potential increase of approximately 45 percentage points

Figure 3. Prevalence-adjusted post-test cancer probabilities of the adnexal mass risk assessment risk groups projected based on the combined validation datasets. The estimated cancer probability bars were superimposed with an interpolation curve by logistic regression.

in cancer probability. Similarly, the estimated LR- of 0.5 for the postmenopausal AMRA LR group, suggests a post-test decrease in cancer probability of approximately 15 percentage points.

As a multivariate index assay, AMRA provided a relatively uniform performance across multiple histological subtypes and performed well for detecting LMPs and stage I/II invasive ovarian cancers. In particular, ROC curve analysis confirmed a statistically significant improvement by AMRA over CA125 overall, as well as in stage I/II ovarian cancer for premenopausal patients for whom the decision of surgery often requires careful consideration.

To aid clinicians in the management of indeterminate masses, key features of AMRA design and actual derivation and implementation included a high sensitivity with its rule-out cutoff resulting a very high NPV for a large portion of the test populations indicated by AMRA as LR. The rule-in cutoff, on the other end, identified a small yet 'case-enriched' group of patients, offering a clinically actionable PPV and at the same time captures a majority of the total cancer cases. Based on the independent validation results from the current study and pending further validation in prospective studies, the AMRA algorithm with its two cutoffs seems to be able to provide the desired

Table 4. Estimated performance metrics of adnexal mass risk assessment risk groups.

AMRA risk group performance metrics [†]	Training set: OVA1 study set				Combined validation sets: OVA500, FHCRC #7788, OVA1-PS1-CO4			
	Pre-menopausal		Post-menopausal		Pre-menopausal		Post-menopausal	
	All Malignancies	Exclude LMP	All Malignancies	Exclude LMP	All Malignancies	Exclude LMP	All Malignancies	Exclude LMP
HR								
Percentage of test population	9.2	9.2	20.6	20.6	7.9	7.9	10.6	10.6
Sensitivity (%)	68.6	72.5	78.4	86.7	66.7	75.9	69.8	75.0
Prevalence, PPV (%)	37.2	30.9	38.2	35.6	42.3	35.8	66.1	60.8
LR+	11.3	11.9	5.6	6.1	13.9	15.9	17.6	18.9
IR								
Percentage of test population	31.4	31.4	47.1	47.1	24.3	24.3	36.7	36.7
Prevalence (%)	2.8	2.2	3.8	2.2	4.0	2.1	6.3	4.8
HR + IR								
Percentage of test population	40.7	40.7	67.6	67.6	32.2	32.2	47.3	47.3
Sensitivity (%)	86.3	90.0	96.6	99.0	85.9	89.7	93.1	95.6
Prevalence, PPV (%)	10.6	8.7	14.3	12.4	13.3	10.3	19.7	17.3
LR+	2.3	2.3	1.5	1.5	2.9	3.0	2.2	2.2
LR								
Percentage of test population	59.3	59.3	32.4	32.4	67.8	67.8	52.7	52.7
Specificity (%)	61.7	61.7	35.6	35.6	70.6	70.6	57.8	57.8
NPV (%)	98.8	98.8	98.9	98.9	99.0	99.0	98.7	98.7
LR-	0.5	0.5	0.6	0.6	0.5	0.5	0.5	0.5

[†] Percentage of test population, post-test prevalence including PPV and NPV were estimated after adjusting pretest cancer prevalence to 5.0% for premenopausal patients and 10.0% for postmenopausal patients.
 AMRA: Adnexal mass risk assessment; HR: High risk; IR: Intermediate risk; LMP: Low-malignant potential tumor; LR: Lower risk; NPV: Negative predictive value; PPV: Positive predictive value.

efficiency in clinical management of patients diagnosed with a suspicious adnexal mass, including recommendation for surgery for HR patients, serial monitoring with ultrasound exam for LR patients, and assessment by clinical impression for IR patients.

Future perspective

There are many reasons a woman with an adnexal mass might wish to avoid immediate surgery if it is safe to do so, including financial cost and the recovery time involved in surgery. However, there is not presently a test on the market specifically aimed at this 'watch and wait' population. The purpose of AMRA is to address this need, so that in the future women with adnexal masses might consider to delay surgery based on their physician's workup and with the added confidence of the test result.

Author contributions

Z Zhang performed design and conceptualization, drafting, data analysis and final approval. RG Bullock performed drafting, data analysis and final approval. H Fritsche performed design and conceptualization, critical editing and final approval. C Dunton performed design and conceptualization and critical editing. L Twiggs performed critical editing.

Financial & competing interests disclosure

This research and manuscript were funded by Vermillion, Inc. Herbert Fritsche is an employee of ASPIRA Labs, a subsidiary of Vermillion, Inc. Rowan G Bullock is an employee of Vermillion, Inc. Zhen Zhang is a consultant with Vermillion, Inc., however contributed to this work in a personal capacity and is independent of his affiliation with Johns Hopkins University. Vermillion, Inc. provided funding for this study. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that all clinical trials referenced in this manuscript underwent IRB approval, and all human participants gave informed consent to have their data used.

Data sharing statement

The authors certify that this manuscript reports the secondary analysis of clinical trial data that have been shared with them, and that the use of this shared data is in accordance with the terms (if any) agreed upon their receipt. The source of this data is: OVA1-001-CO1, OVA1-001-CO3, OVA1-7788, OVA1-PS1-CO4. As the data are proprietary, no individual clinical trial data will be shared other than the statistical summaries already available.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Summary points

- Adnexal mass risk assessment (AMRA) is a multivariate index assay of serum biomarkers developed to stratify women with adnexal masses into three risk groups for the purpose of capturing the relatively small number of malignancies in a high-risk (HR) group with a high positive predictive value, and keeping a low-risk (LR) group with a very high negative predictive value. The high confidence in the AMRA LR result could allow the doctor to include it in a 'watch and wait' tool as a part of the workup for an adnexal mass.
- The algorithm was developed using retrospective data from prospectively collected samples in previous clinical trials. All serum samples were from women with confirmed adnexal masses who had been scheduled for surgery.
- Two separate predictive models (algorithms) were derived for the pre- and post-menopausal patient populations, respectively. For each algorithm, the training dataset was also used to determine a 'rule-in' cutoff to identify HR group and a 'rule-out' cutoff to identify relatively LR group. The samples between the two cutoffs were classified as intermediate risk (IR).
- To project the performance of the AMRA algorithms in its intended population, we assumed a pretest prevalence of 5 and 10% in pre- and post-menopausal test populations, respectively, and estimated clinical performance was calculated.
- AMRA demonstrated sharp differentiation of post-test cancer probabilities across the three risk groups that should be clinically actionable. For both pre- and post-menopausal population, the relatively small HR group captured over two-third of the total cancer cases. When low-malignant potential tumors were excluded, the sensitivity of HR both increased to 75% or above.
- The LR group identified a significant portion of the test population (67.8 and 52.7% for pre- and post-menopausal, respectively) with an negative predictive value close to 99%.
- The sensitivity and positive predictive value of the combined group of HR and IR to be at 85.9 and 13.3%, respectively, for premenopausal patients and 93.1 and 19.7%, respectively, for postmenopausal patients.
- As a multivariate index assay, AMRA provided a relatively uniform performance across multiple histological subtypes and performed well for detecting low-malignant potential tumors and stage I/II invasive ovarian cancers with significant improvement over CA125.
- Based on the independent validation results from the current study and pending further validation in prospective studies, the AMRA algorithm with its two cutoffs seemed to be able to provide the desired efficiency in clinical management of patients diagnosed with a suspicious adnexal mass, including recommendation for surgery for HR patients, serial monitoring with ultrasound exam for LR patients and assessment by clinical impression for IR patients.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Demir R, Marchand G. Adnexal masses suspected to be benign treated with laparoscopy. *J. Soc. Laparoendosc. Surg.* 16(1), 71–84 (2012).
2. Farghaly S. Current diagnosis and management of ovarian cysts. *Clin. Exp. Obstet. Gynecol.* 41(6), 609–612 (2014).
3. Sadowski E, Paroder V, Patel-Lippmann K *et al.* Indeterminate adnexal cysts at US: prevalence and characteristics of ovarian cancer. *Radiology* 287(3), 1041–1049 (2018).
4. Modesitt S, Pavlik E, Ueland F *et al.* Risk of malignancy in unilocular ovarian cystic tumors less than 10 centimeters in diameter. *Obstet. Gynecol.* 102(3), 594–599 (2003).
5. Ueland F, Desimone C, Seamon L *et al.* Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obstet. Gynecol.* 117, 1289–1297 (2011).

6. Bristow R, Smith A, Zhang Z *et al.* Ovarian malignancy risk stratification of the adnexal mass using a multivariate index assay. *Gynecol. Oncol.* 128, 252–259 (2013).
 - **Validated multivariate index assay (MIA) on 494 subjects, finding a sensitivity of 95.7% overall, and an early stage sensitivity of 91.4%.**
7. Coleman R, Herzog T, Chan D *et al.* Validation of a second-generation multivariate index assay for malignancy risk of adnexal masses. *Am. J. Obstet. Gynecol.* 215(1), 82.e1–82.e11 (2016).
 - **Validated multivariate index assay second generation on 519 subjects, finding an overall sensitivity of 91.3% and a specificity of 69.1%, a significant improvement over MIA.**
8. Molinaro A. Diagnostic tests: how to estimate the positive predictive value. *Neuro Oncol. Practice* 2(4), 162–166 (2015).
 - **Discusses the ways to calculate positive predictive values (PPV), and the clinical significance of PPV, namely that PPV is prevalence-dependent cannot be calculated from a population with a different disease prevalence.**
9. McGee S. Simplifying likelihood ratios. *J. Gen. Intern. Med.* 17(8), 646–649 (2002).
 - **Explains the clinical utility of likelihood ratios and shows how they can be applied independent of disease prevalence.**